Ereignisdatenanalyse - Beispiele, Probleme und Perspektiven

Diekmann, Andreas

Ereignisdatenanalyse - Beispiele, Probleme und Perspektiven

Ereignisdaten sind spezielle zeitbezogene Daten, die über die Abfolge von Ereignissen und die Länge der Zeitintervalle zwischen den untersuchten Ereignissen informieren. Ein Beispiel ist die Berufsbiographie einer Person mit den Ereignissen Berufseintritt, Wechsel des ersten Berufs, Beginn einer Arbeitslosigkeitsphase, Wiederbeschäftigung, erneuter Berufswechsel, Pensionierung. Das Ausgangsmaterial der Datenanalyse ist die Zeit zwischen zwei Ereignissen, z.B. die Zeit vom Beginn der Arbeitslosigkeit bis zur Wiederbeschäftigung. Die Länge einer solchen "Episode" wird als "Wartezeit", "Verweildauer" oder "Ankunftszeit" bezeichnet. Ereignisdaten geben somit Auskunft über den Ablauf sozialer Prozesse. Sie sind wesentlich informativer als Querschnitts- und sogar informativer als Panel-Daten, die nur Momentaufnahmen zu den Erhebungszeitpunkten eines Panels darstellen. Allerdings können mit Panel-Designs und auch mit Querschnittserhebungen vergangene Abläufe und damit Ereignisdaten retrospektiv erhoben werden.

Der vorliegende Artikel referiert zunächst im Überblick die Grundideen der Ereignisdatenanalyse anhand verschiedener Beispiele aus der empirischen Sozialforschung. Anschließend wird auf den Zusammenhang zwischen den in der Ereignisdatenanalyse gebräuchlichen Modellen und den Modellen zur Erforschung sozialer Diffusion aufmerksam gemacht. Einige Forschungsperspektiven werden im Schlußabschnitt aufgezeigt.

1. Von der Erprobung zur Anwendung

Zur Analyse von Ankunftszeiten werden spezifische Verfahren benötigt, die heute in der Soziologie unter dem Oberbegriff Ereignisdaten - oder auch Verlaufsdatenanalyse, in der Statistik zumeist unter dem Begriff "Survivalanalyse" - zusammengefaßt werden. Waren die in den siebziger Jahren breiter angewandten methodischen Neuerungen (Pfadanalyse, LISREL, log-lineare Tabellenanalyse) vorwiegend, wenn auch nicht ausschließlich "statisch" ausgerichtet und auf Querschnittsdaten bezogen, so ist die Ereignisdatenanalyse explizit "dynamisch" und prozeßorientiert. Von frühen Vorläufern abgesehen, sind Anwendungen in der Soziologie in stärkerem und zunehmendem Maße seit etwa zehn Jahren zu konstatieren. Während die Standard-Lehrbücher zur Survivalanalyse (vgl. die "Wiley-Series in Probability and Mathematical Statistics": Elandt-Johnson/Johnson 1980; Kalbfleisch/Prentice 1980; Lawless 1982; Nelson 1982) noch ausschließlich auf biometrische und demographische Anwendungen sowie Reliabilitätsuntersuchungen in der Technik ausgerichtet waren, sind um die Mitte der achtziger Jahre mehrere Lehrbücher und Aufsatzsammlungen mit im engeren Sinne soziologischem Bezug erschienen (Coleman 1981; Tuma/Hannan 1984; Diekmann/Mitter 1984 a,b; Allison 1984; Andreß 1985; Blossfeld/Hamerle/Mayer 1986; Mayer/Tuma 1987). Diese dokumentieren, ebenso wie eine wachsende Zahl von Zeitschriftenveröffentlichungen, die breite Rezeption der Ereignisdatenanalyse in der empirisch forschenden Soziologie. Ähnliches gilt für die empirische Mikroökonomie. Anwendungen und Weiterentwicklungen der Ereignisdatenanalyse sind hier vor allem mit den Namen Heckman und Singer (1982, 1984 a,b) verbunden.

In der Bundesrepublik sind mittlerweile zahlreiche empirische Forschungsprojekte mit der Erhebung und Auswertung von Ereignisdaten befaßt. Dazu zählen die Lebensverlaufsstudie des SFB 3 von K.U. Mayer, das "Sozioökonomische Panel", das in der Erhebungsphase befindliche DJI-Projekt "Dauerbeobachtung von Familien", das Kölner Gymnasiastenpanel, die Studien über Berufsverläufe und die Lebensdauer von Unternehmen des Projekts Ziegler im Münchner SFB 333 und die geplanten Arbeiten über "Statuspassagen" im neugegründeten sozialwissenschaftlichen SFB an der Universität Bremen, um nur einige der größeren Projekte zu nennen. Fast alle diese und weitere Projekte mit ereignisanalytischen Untersuchungsverfahren wurden oder werden von ZUMA beraten. Auch dies ist ein Grund für ZUMA, sich in Zukunft verstärkt Anwendungsproblemen und der Weiterentwicklung ereignisanalytischer Methoden in der Forschung, Beratung und Weiterbildung zu widmen.

Das große Interesse, das die Ereignisdatenanalyse als sicher nicht einziges Instrument bei der Untersuchung des sozialen Wandels in der Soziologie der achtziger Jahre gefunden hat, relativiert die Sprechweise von einer "neuen" Methode. Mittlerweile wurden in der praktischen Forschung zahlreiche Erfahrungen gewonnen. Die Ereignisdatenanalyse ist damit aus dem Erprobungsstadium hinaus in das Stadium "normaler" Anwendungen in der soziologischen Forschung gelangt.

2. Ausgangspunkte der Ereignisdatenanalyse: Daten und Zensierungsproblem

Das Ausgangsmaterial der Ereignisdatenanalyse sind - wie bereits erwähnt - spezifische, besonders informative zeitbezogene Daten, die über die Ankunftszeit bis zum Eintreffen eines Ereignisses Aufschluß geben. Es handelt sich um einen Spezialfall von Längsschnittdaten, nicht aber um Zeitreihendaten im üblichen Sinne. Während letztere, meist in aggregierter Form, für eine oder wenige Untersuchungseinheiten vorliegen, sind Ereignisdaten Individualdaten, die über die individuellen Verläufe bei einer Vielzahl von Untersuchungseinheiten informieren. So weist z.B. die zwischen 1950 und 1988 jährlich ermittelte Arbeitslosenquote in der Bundesrepublik die Struktur einer Zeitreihe auf. Wird jedoch bei einer Stichprobe von 1000 Personen, die im Jahre 1985 ihren Arbeitsplatz verloren haben, jeweils die Dauer der Arbeitslosigkeit erhoben, so handelt es sich hierbei um eine völlig andere Datenstruktur - nämlich um Ereignisdaten.

Die Einheiten der Analyse sind nicht notwendigerweise Personen. Als Untersuchungseinheiten können zum einen Nationen, Organisationen etc. betrachtet werden, etwa wenn die Lebensdauer neugegründeter Firmen zum Gegenstand der Untersuchung gemacht wird. Zum anderen kann eine Person bei wiederholbaren Ereignissen mehrere Episoden - z.B. mehrere Arbeitslosigkeitsintervalle - aufweisen. Bei sogenannten "all-spell"-Analysen ist in diesem Fall nicht die Person, sondern jede Episode eine Untersuchungseinheit. Dabei tritt allerdings häufig das Problem der statistischen Abhängigkeit der Episodenlängen auf.

Substantielle Anwendungsbeispiele lassen sich in nahezu allen Bereichen der Sozialwissenschaft finden. Zu nennen sind in der Bevölkerungssoziologie etwa Analysen des Heiratsalters, der Ehedauer, von Geburtenintervallen und der Sterblichkeit; in der Kriminologie Analysen von Rückfallzeiten nach der Haftentlassung (z.B. beim Vergleich von Alternativen zur Freiheitsstrafe mit dem herkömmlichen Strafvollzug); in der Migrationsforschung Analysen der Wohndauer; in der Mobilitätsforschung und Soziologie des Arbeitsmarktes die Verweildauer in einer beruflichen Tätigkeit, die Dauer der Arbeitslosigkeit, die Betriebszugehörigkeitsdauer oder die Lebensdauer von Unternehmen; in der Sozialpsychologie die Untersuchung von Reaktionszeiten; in der Interviewforschung die Analyse von Fragebogen-Rücklaufzeiten usw.

Für die Ereignisdatenanalyse charakteristisch ist, daß diese Verfahren konsistente Schätzungen der Parameter der Ankunftszeitenverteilung auch dann erlauben, wenn sogenannte zensierte Daten vorliegen.

Da in der Praxis häufig nur ein Teil der Untersuchungseinheiten im Beobachtungszeitraum ein Ereignis aufweist, kann nicht bei allen Fällen die vollständige Ankunftszeit ermittelt werden. Bei einer Untersuchung der Ehedauer werden nicht alle Ehen im Beobachtungszeitraum geschieden; bei einer Studie zur Arbeitslosigkeitsdauer nicht alle Arbeitslose im Beobachtungszeitraum wieder in Lohn und Brot stehen. Bekannt ist bei solchen (rechts-)zensierten Daten nur eine untere Schranke der tatsächlichen Ankunftszeit. Herkömmliche Analyseverfahren, wie z.B. die Pfad-, Regressions- oder Varianzanalyse, sind bei Präsenz zensierter Daten nicht einsetzbar. Wie man sich leicht vergegenwärtigen kann, sind beim Vorliegen zensierter Daten nicht einmal einfache, zentrale Kenngrößen der Ankunftszeitenverteilung, wie der Mittelwert oder Median, auf gewohnte Weise schätzbar. Die Ignorierung zensierter Daten als "missing values" führt eventuell je nach Datenlage zu einer krassen Unterschätzung der mittleren Ankunftszeit, wie folgendes Beispiel zeigt.

Bei der Berechnung des mittleren Heiratsalters nach Geburtskohorten (vgl. Tabelle 1) wird das Alter der (noch) ledigen Personen als zensiertes Datum berücksichtigt. Ignoriert man die ledigen Personen einfach als "missing values", so wird nur die Gruppe der besonders rasch in den Ehestand eintretenden Mitglieder einer Geburtshohorte erfaßt.

Wie anhand der Tabelle zu erkennen ist, wird der Median mit steigendem Zensierungsanteil stark unterschätzt. Ohne die Berücksichtigung der zensierten Fälle würde man fälschlicherweise schlußfolgern, daß das Heiratsalter in diesem Jahrhundert stetig gefallen ist. Tatsächlich aber ist die Entwicklung U-förmig. Bei Kohortenanalysen sind hohe Zensierungsquoten zwangsläufig in den jüngeren Kohorten anzutreffen. Gerade diese Kohorten sind aber von besonderem Interesse für die Abschätzung künftiger Entwicklungen.

Tabelle 1: Median-Ersthelratsalter nach Kohorten und Geschlecht

Geburts-kohorte	"Sterbetafel"-Schätz-methode bei Berücksichtigung zensierter Daten		Median-Schätzung ohne Berücksichtigung zensierter Zeiten (nur Verheiratete)	Zensierungs-anteil in %
	Frauen	Männer	Männer	(Ledigenquote)
1956-66	24.4	27.9	22.5	83%
1946-55	22.2	25.8	24.6	24%
1936-45	22.9	25.2	25.1	4%
1926-35	23.6	26.0	25.9	2%
vor 1926	24.7	28.2	28.1	2%

Daten: ALLBUS 1982 und 1984 (Diekmann 1987: 72,132).

Zur Schätzung der Verteilungsparameter der Ankunftszeitenverteilung unter Berücksichtigung zensierter Daten und insbesondere auch zur Schätzung der Einflußstärke unabhängiger Variablen auf den Verlauf eines untersuchten sozialen Prozesses, wird in der Ereignisdatenanalyse ein Modell des stochastischen Prozesses spezifiziert, von dem angenommen wird, daß dieser die beobachtete Verteilung der Ankunftszeiten generiert hat. Die Parameter können sodann mittels der Maximum-Likelihood- oder Partial-Likelihood-Methode geschätzt werden. Programmpakete wie BMDP, SAS, RATE, GLIM, GAUSS, LIMDEP und weitere Spezialprogramme liefern hierzu die nötige Softwareunterstützung.

3. Modelle und Parameterschätzung

3.1 Übergangsrate und Ankunftszeitenverteilung

Definiert man Ereignisse als Wechsel der Ausprägung einer qualitativen Zustandsvariable Y(t) und können Ereignisse zu beliebigen Zeitpunkten t auftreten, so empfehlen sich zur Modellierung stochastische Prozeß-Modelle mit diskreter Zustandsvariable und stetiger Zeit. Bei vielen Anwendungen genügt es, einen Ausgangszustand und einen Zielzustand zu betrachten. Die Zustandsvariable Y(t) hat in diesem Fall nur zwei Ausprägungen: bei der Untersuchung des Heiratsalters z.B. die Ausprägung "0" für "ledig" und die Ausprägung "1" für "Erstheirat". Ein Ereignisdatum ist in diesem Kontext ein Informationstripel, das den Ausgangszustand, den Zielzustand und die Verweildauer im Ausgangszustand registriert. Komplexe Mehr-Zustands-Modelle lassen sich häufig (bei Markov- oder Semi-Markov-Prozessen mit unkorrelierten Risiken) für jede Ausgangs- und Zielzustandskombination in eine entsprechende Anzahl von Zwei-Zustands-Modellen zerlegen, die jeweils separat an Daten schätzbar sind.

Das Ziel der Modellbildung besteht in der mathematischen Ableitung der Ankunftszeitenverteilung der nicht-negativen Zufallsvariablen T aus dem zugrunde gelegten Modell. Ist die Ankunftszeitenverteilung bekannt, so können deren Parameter mit der Maximum-Likelihood-Methode anhand von Ereignisdateh geschätzt und verschiedene Modellimplikationen geprüft werden. Das an Daten geschätzte und geprüfte Modell des stochastischen Prozesses ist dann u.a. für prognostische Zwecke verwendbar.

Was heißt nun Modellspezifikation? Das erste Glied in der Kette ist hier im allgemeinen die Übergangsrate (Hazardrate, Risiko) r(t), die eng mit der bedingten Wahrscheinlichkeit eines Zustandswechsels zusammenhängt. Die Übergangsrate wird definiert als Grenzwert der (bedingten) Übergangswahrscheinlichkeit $q(t,t+\Delta t)$ dividiert durch das Zeitintervall Δt

(1) $$r(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \; q(t,t+\Delta t)$$
$$= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \; P(t+\Delta t > T \geq t \mid T \geq t).$$

Wird die kumulierte Verteilungsfunktion der Ankunftszeit mit

(2) $F(t) = P(T \leq t)$ bezeichnet,

die (unbedingte) Wahrscheinlichkeitsdichtefunktion mit

(3) $$f(t) = \frac{dF(t)}{dt} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \; P(t+\Delta t > T \geq t)$$

und das Komplement zu F(t), die Überlebensfunktion, mit

(4) $G(t) = P(T \geq t) = 1-F(t)$,

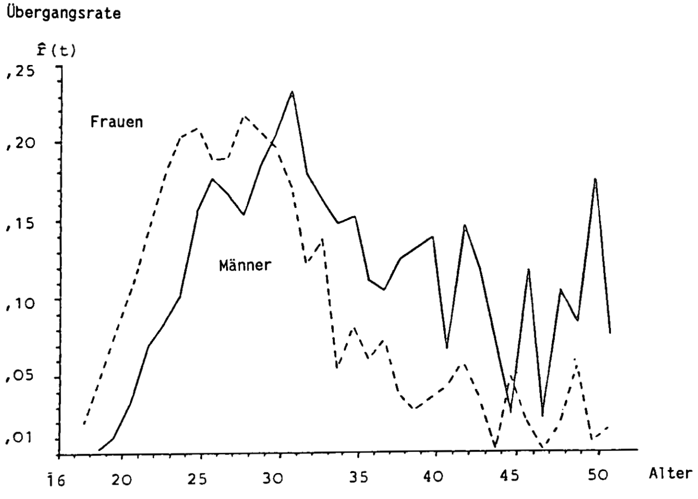
dann folgt aus der Definition der Übergangsrate (1) die grundlegende Beziehung

(5) $$r(t) = \frac{f(t)}{G(t)} .$$

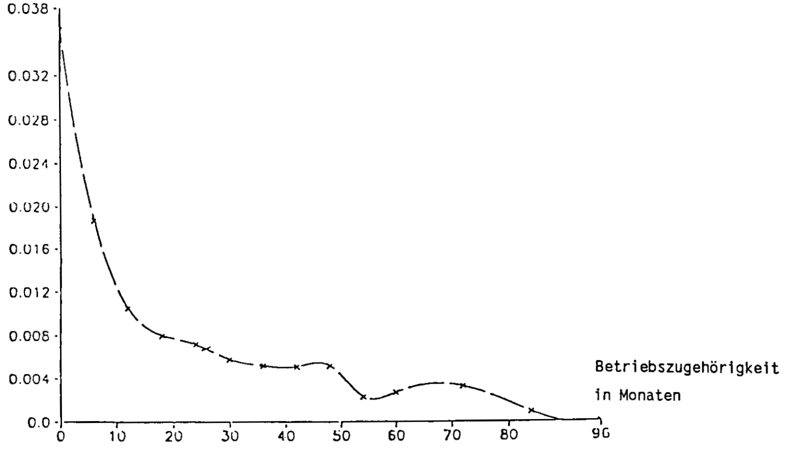
Anschaulich gesprochen informiert F(t) über den Anteil von Personen bzw. Untersuchungseinheiten, die bis zum Zeitpunkt t ein Ereignis aufweisen. Die Überlebensfunktion G(t) informiert über den Anteil der Personen, die bis zum Zeitpunkt t im Ausgangszustand verbleiben, und r(t) gibt näherungsweise über die Wahrscheinlichkeit Auskunft, daß in der auf t folgenden kurzen Zeiteinheit ein Ereignis auftritt, vorausgesetzt bis t ist kein Zustandswechsel erfolgt.

Abbildung 1 zeigt sogenannte nicht-parametrische Sterbetafel-Schätzungen der Übergangsrate für die Ereignisse "Erstheirat" und "Betriebswechsel". In beiden Fällen ist die Übergangsrate zeitabhängig: Sie weist ein nicht-monotones

Abbildung 1: Lifetable-Schätzung der Übergangsrate für Erstheirat (oben) und das Ausscheiden aus einem Betrieb (unten)



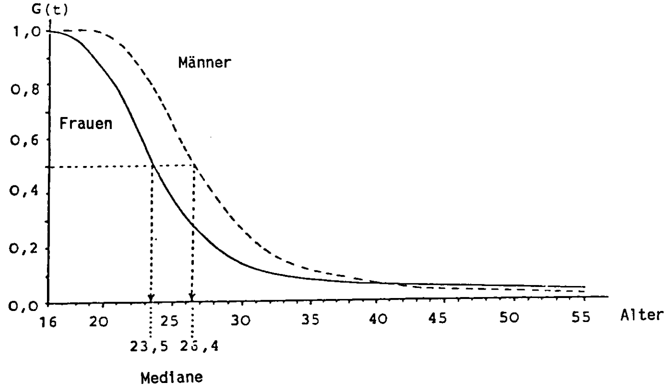
N=2961 Frauen, 2593 Männer, Allbus 1982 und 1984, Diekmann 1987:105



N=1432 deutsche Arbeitnehmer aus einem metallverarbeitenden Betrieb ("Südwerk"), Diekmann/Preisendörfer 1988

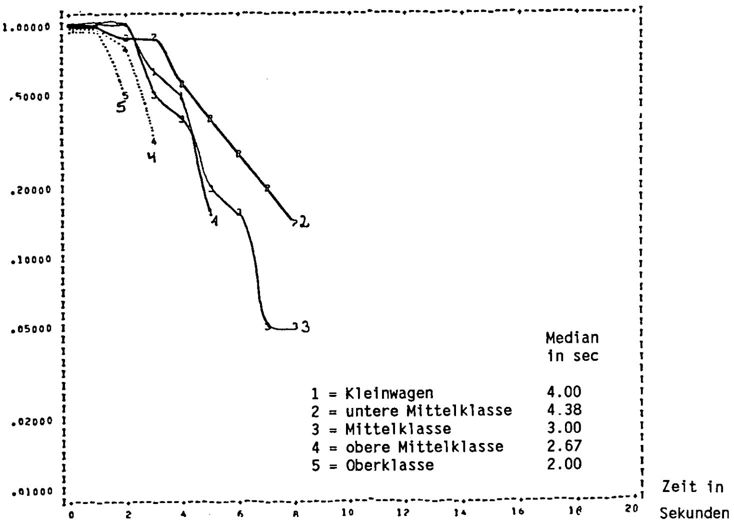
Abbildung 2: Beispiele für "Überlebenskurven": Ersteheirat (oben) und Feldexperiment "Aggression im Straßenverkehr" (unten)

Ledigenquoten



Allbus-Daten, vgl. Abbildung 1

Logarithmierte "Überlebensfunktion"



N=57, Heinritz/Jungbauer-Gans/Krassnig 1986 "glockenförmiges" Muster bei den Heiratsdaten und einen monoton sinkenden Verlauf bei den Betriebswechseldaten auf.

Die geschätzten "Überlebenskurven" (die Ledigenquoten) nach Geschlecht für die Heiratsdaten gehen aus Abbildung 2 (oben) hervor. Bei geringer Fallzahl sind die geschätzten Kurven weniger "glatt", wie das Untersuchungsbeispiel "Aggression im Straßenverkehr" (Abbildung 2 unten) verdeutlicht. Bei diesem Feldexperiment wurde die Weiterfahrt eines nachfolgenden Fahrzeugs nach dem Umschalten der Ampel auf "grün" durch ein Versuchsfahrzeug blockiert. Dann wurde die Zeit bis zum Betätigen der Hupe bei dem blockierten Fahrzeug erhoben. Die Abbildung läßt erkennen, daß die "Überlebenskurven" mit der Größe der Autoklasse nach "links" rücken. Je prestigeträchtiger das blockierte Auto, desto kürzer ist die Zeit bis zum Betätigen der Hupe (vgl. auch die Median-Reaktionszeit in Abbildung 2).

Um die Verteilung der Ankunftszeiten abzuleiten, wird Ausdruck (5) mit Hilfe von (3) und (4) folgendermaßen umgeformt:

$$(6) \qquad r(t) = \frac{\frac{dF(t)}{dt}}{1 - F(t)}$$

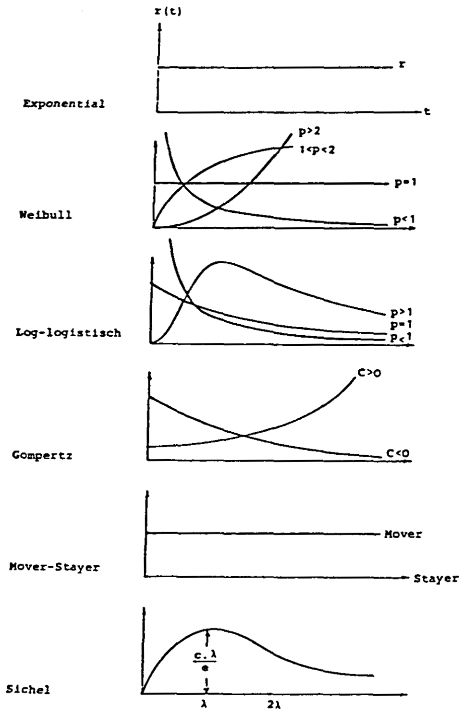
Die Lösung der Differentialgleichung (6) ist die gesuchte Beziehung zwischen der Übergangsrate und der Ankunftszeitenverteilung:

$$(7) \qquad F(t) = 1 - \exp\left(-\int_0^t r(\tau)\, d\tau\right) .$$

Die Chance oder das Risiko eines Wechsels, ausgedrückt durch die Übergangsrate $r(t)$, kann sowohl von der Zeit t, von der Verweildauer im Ausgangszustand, als auch von exogenen Einflüssen (Kovariaten) abhängen. So wird etwa bei Arbeitslosen die Chance, einen neuen Arbeitsplatz zu erhalten, mit der Dauer der Arbeitslosigkeit abnehmen (negative Verweildauerabhängigkeit), aber darüber hinaus auch mit der Qualifikation, der Arbeitsmarktlage, dem Geschlecht und weiteren Merkmalen variieren.

Anhand von Ausdruck (7) ist erkennbar, daß die Form der Ankunftszeitenverteilung $F(t)$ eindeutig durch die jeweilige Zeitabhängigkeit der Übergangsrate bestimmt wird. Je nachdem, ob die Übergangsrate zeitkonstant ist (Poisson-Prozeß), mit der Zeit ansteigt, absinkt oder nicht-monoton (z.B. umgekehrt U-förmig) verläuft, wird man verschiedene Verteilungen der Ankunftszeiten erhalten. Für eine Reihe von Standardsituationen kann dabei auf vorliegende Modelle zurückgegriffen werden, die auch softwaremäßig realisiert sind (Exponential-, Gompertz-, Weibull-, log-logistisches-, Sichel-Modell etc., vgl. Abbildung 3).

Abbildung 3: Parametrische Übergangsratenmodelle



Aus: Diekmann/Mitter, 1984a: 150

3.2 Das Basis-Modell mit konstanter Übergangsrate

Im einfachsten Fall ist die Übergangsrate zeitunabhängig mit $r(t)=r$. Es folgt dann aus (7), (3) und (4) die Exponentialverteilung mit

$$F(t) = 1 - e^{-rt}$$

$$f(t) = re^{-rt}$$

$$G(t) = e^{-rt} .$$

Der Mittelwert der Ankunftszeit ist bei diesem Modell der reziproke Wert der Rate

$$\bar{T} = 1/r$$

und für den Median gilt

$$\text{Median } (T) = \ln 2/r = 0.69 \cdot \bar{T} .$$

Wie bei rechtsschiefen Verteilungen zu erwarten, befindet sich der Median immer vor dem Mittelwert. Alle wichtigen Charakteristika des Prozesses, die Ankunftszeitenverteilung, die Überlebensfunktion, die mittlere Ankunftszeit und der Median oder die "Halbwertszeit" werden eindeutig durch die Übergangsrate festgelegt. Ist die Übergangsrate ferner von Kovariaten abhängig, so sind die erwähnten Charakteristika des Prozesses für beliebige Ausprägungskombinationen von Kovariaten prognostizierbar.

3.3 Einführung von unabhängigen Variablen (Kovariaten)

Da die Übergangsrate gemäß Definition (1) eine nicht-negative Größe ist, empfiehlt sich in der Regel - demonstriert am Beispiel des Exponentialmodells - die folgende log-lineare Spezifikation der Übergangsrate in Abhängigkeit von Kovariaten

$$r = \exp (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_m x_m)$$

mit den empirisch zu schätzenden β-Parametern und den Kovariaten x_1, x_2,... x_m. Schreibt man für $\exp\beta = \alpha$, so ergibt sich der Ausdruck:

$$r = \alpha_0 \, \alpha_1^{x_1} \, \alpha_2^{x_2} \ldots \alpha_m^{x_m} .$$

Die α-Effekte sind hierbei relativ anschaulich als Prozent-Effekte der Kovariate auf die Rate $((\alpha-1) \cdot 100)$ und (nur) beim Exponentialmodell wegen der Beziehung $\bar{T} = 1/r$ auch als Prozent-Effekte auf die mittlere Ankunftszeit und den Median $((1/\alpha-1) \cdot 100)$ zu interpretieren. Ergibt sich demnach ein Effekt von z.B. 1.04, so bedeutet dies, daß bei einer Erhöhung der unabhängigen Variable um eine Einheit ein Anstieg der Rate von 4% und bei Zugrundelegung des Exponentialmodells eine Verringerung der mittleren Ankunftszeit bzw. des Medians um 3.8% zu erwarten ist.

Bei den Betriebswechseldaten zeigte sich, daß das Wechselrisiko mit der Verweildauer im Betrieb absinkt (vgl. Abbildung 1 unten). Das Exponentialmodell mit zeitkonstanter Übergangsrate ist in diesem Fall nicht angemessen. Dagegen erlaubt das Gompertz- und das semi-parametrische Cox-Modell (dazu der folgende Abschnitt) die Modellierung monoton sinkender Übergangsraten. Auch hier empfiehlt sich eine log-lineare Spezifikation für die Modellparameter in Abhängigkeit von den Kovariaten (Tabelle 2).

Tabelle 2: Multivariate Analyse einiger Einflußfaktoren auf die Übergangsrate des Ausscheidens aus einer Firma

	Cox-Modell**	Gompertz-Modell**
Geschlecht	1.991* (4.145)	2.073* (4.386)
Alter bei Eintritt in die Firma	0.949* (9.050)	0.948* (9.196)
Anfangs-Lohngruppe	0.742* (7.213)	0.728* (7.648)
Lohngruppenanstieg	0.510* (2.857)	0.489* (3.078)
Beschäftigungsgruppe Produktionsarbeiter	0.887 (0.979)	0.874 (1.097)
Eintritt in Kontraktionsperiode Ende 1981-84	1.116 (0.827)	1.102 (0.739)
Konstante α_o	------	0.391* (3.405)
Konstante c_o	------	-0.042* (15.484)
Chi^2(df)	181.36 (6)	616.74 (7)
% zensierte Fälle	54.6	54.6
N	1423	1423

Referenzgruppe für dichotome Variablen: Frauen, Beschäftigung in der Qualitätskontrolle, Eintritt in der Expansionsperiode 1976-81

* Signifikant für p=0,05; t-Werte in Klammern.

** Cox-Modell: $r(t) = h_o(t)\ \alpha_1^{x_1}\ \alpha_2^{x_2} \ldots \alpha_m^{x_m}$

Gompertz-Modell: $r(t) = \alpha_o\ \alpha_1^{x_1} \ldots \alpha_m^{x_m}\ e^{c_o t}$

Die α-Koeffizienten sind als %-Effekte auf die Übergangsrate zu interpretieren. Ein Koeffizient z.B. von 0.949 für das Alter bei Eintritt (Cox-Modell) besagt, daß bei einem um ein Jahr höheren Lebensalter bei Eintritt in die Firma das Risiko des Betriebswechsels um (1-0.949)x100 = 5% absinkt.

Untersucht wurde der Einfluß verschiedener unabhängiger Variablen auf das Risiko bzw. die Übergangsrate des Ausscheidens aus dem Betrieb (Daten siehe Abbildung 1 unten). Die Ergebnisse der Schätzung gehen aus Tabelle 2 hervor. Es zeigt sich, daß das Risiko signifikant am höchsten ist in der Gruppe der männlichen Arbeitnehmer mit geringem Eintrittsalter, geringem Anfangslohn und geringem Lohnanstieg. Die Studie belegt die These, daß steigende Lohnprofile das Risiko eines Betriebswechsels bremsen können. Der signifikant negative Wert der Konstante c_0 des Gompertz-Modells macht deutlich, daß die Übergangsrate auch bei multivariater Analyse mit der Betriebszugehörigkeitsdauer absinkt.

3.4 Schätzung der Parameter

Gemäß Ausdruck (7) sowie (3) und (4) treten die α- bzw. β-Parameter bei obiger log-linearer Spezifikation der Ratengleichung unmittelbar in der kumulierten Ankunftszeitenverteilung F(t), der Dichteverteilung f(t) und der Überlebensfunktion G(t) auf. Anhand der beobachteten Ankunftszeiten und Kovariatenwerte können die Parameter mit der Maximum-Likelihood-Methode unter Einschluß der zensierten Beobachtungen geschätzt werden. Maximiert in bezug auf die α- bzw. β-Parameter wird das Produkt über die Likelihoodbeiträge der Untersuchungseinheiten i=1,...,N mit dem Likelihoodbeitrag der Person i:

$$L_i = f(t_i, \underline{x}_i, \underline{\beta})^{d_i} \, G(t_i, \underline{x}_i, \underline{\beta})^{(1-d_i)}$$

Dabei bezeichnet $\underline{x}_i$ den Vektor der Kovariatenwerte, $\underline{\beta}$ den Vektor der Parameter und d_i eine Indikatorvariable mit dem Wert d_i=1 bei einer nicht-zensierten und d_i=0 bei einer zensierten Beobachtung. Der Typ der Verteilung f(.) und der zugehörigen Überlebensfunktion G(.) wird gemäß (7) durch die Wahl der Verweildauerabhängigkeit der Übergangsrate festgelegt.

Die Auswahl eines Modells für eine konkrete empirische Analyse wird sich sowohl an theoretischen Gesichtspunkten orientieren, als auch an bisherige empirische Studien anknüpfen. Die nicht-parametrischen Sterbetafel-Schätzungen informieren zudem über den Verlauf der Übergangsrate in Abhängigkeit von der Verweildauer. So wurde z.B. bei Analysen des Heiratsalters häufig das log-logistische Modell verwendet, weil es u.a. die Modellierung einer umgekehrt U-förmigen Ratenfunktion erlaubt. Dieses Bild ergibt sich aus nichtparametrischen Schätzungen der Übergangsrate (Abbildung 1 oben).

3.5 Semi-parametrische Cox-Regression

Neben der oben erläuterten voll-parametrischen Schätzung (die Übergangsrate wird hier als parametrische Funktion der Verweildauer im Ausgangszustand und der Kovariate spezifiziert) und der weitgehend voraussetzungslosen nichtparametrischen Sterbetafel-Schätzung (im Englischen optimistischer "life-table", zur Einführung vgl. Diekmann/Mitter 1984a: Kap. 3) steht als dritte Methode die semi-parametrische Cox-Regression zur Verfügung. Die Übergangsrate wird hierbei als parametrische Funktion der Kovariate spezifiziert, der Typ der Verweildauerabhängigkeit hingegen offen gelassen. Die Ratenfunktion hat die Form:

$$r(t) = h_0(t) \exp(\underline{B}'\underline{x})$$

mit einer beliebigen Basis-Hazardratenfunktion $h_0(t)$. Das Modell erlaubt einerseits wegen der Allgemeinheit der Basis-Hazardratenfunktion eine flexible Modellierung von Ratenverläufen, unterstellt aber andererseits die (prüfbare) Annahme der Proportionalität der Risiken. Da $h_0(t)$ nicht als Funktion spezifiziert wird, ist es auch nicht möglich, die volle Likelihood-Funktion zu formulieren. Mittels der Partial-Likelihood-Methode sind jedoch Schätzungen der B-Parameter erzielbar (vgl. das Beispiel in Tabelle 2), deren "Qualität" an Maximum-Likelihood-Schätzungen heranreichen. Die Arbeiten des britischen Statistikers Cox (lehrbuchartig präsentiert vor allem in Kalbfleisch/Prentice 1980) haben wegen der Flexibilität des Modells der Survivalanalyse und ihrer breiten Nutzung erhebliche Impulse gegeben. Allerdings ist zu berücksichtigen, daß die Proportionalitätsannahme nicht immer erfüllt sein muß, und voll-parametrische Modelle wegen ihres höheren Informationsgehalts und der einfacheren Handhabbarkeit für prognostische Zwecke gewisse Vorzüge aufweisen können.

Für inferenzstatistische Zwecke stehen bei parametrischen, semi-parametrischen und nicht-parametrischen Verfahren verschiedene Tests auf Signifikanz der geschätzten Parameter und die Unterschiedlichkeit von Überlebenskurven zur Verfügung. In einer Reihe wichtiger Fälle gelingt es, exakte Tests zu formulieren. Im allgemeinen aber werden asymptotische Tests verwendet, die sich aus den "Large-sample"-Theoremen der Likelihood-Theorie ergeben.

4. Prozesse sozialer Diffusion und Ereignisdatenanalyse

Modelle sozialer Diffusion haben in der mathematischen Soziologie eine lange Tradition. Die kumulative Ausbreitung von Neuerungen, die Übernahme neuer Techniken, die Verbreitung von Informationen sowie eine Vielzahl sozialer Prozesse, die den Gesetzen der Imitation folgen, können formal durch Modelle sozialer Diffusion abgebildet werden (als Überblick: Mahajan/Peterson 1985; Hamblin/Jacobsen/Miller 1973). Weniger bekannt ist hingegen, daß diese Modelle einen engen Zusammenhang zu den Modellen der Ereignisdatenanalyse aufweisen. Das log-logistische Modell in der Ereignisdatenanalyse ist sogar aus der Gleichung eines Diffusionsmodells herleitbar (Diekmann 1988) und damit tiefer begründbar. Der Zusammenhang zwischen beiden Forschungsrichtungen ist in zweierlei Hinsicht von Bedeutung: Zum einen kann die Schätzung der Parameter von Prozessen sozialer Diffusion mit den Techniken der Ereignisdatenanalyse verbessert werden, zum anderen wird das Spektrum der in der Ereignisdatenanalyse gebräuchlichen Modelle durch die Einbeziehung von Diffusionsmodellen erweitert. Wie sich leicht zeigen läßt, sind Diffusionsmodelle als Modelle für Ankunftszeiten mit jeweils spezifischen Übergangsratenfunktionen interpretierbar.

Betrachten wir die "klassische" Studie von Coleman/Katz/Menzel (1957), die für "isolierte" Ärzte einerseits und in ihr soziales Netzwerk "integrierte" Ärzte auf der anderen Seite jeweils unterschiedliche Diffusionsprozesse für die erstmalige Verschreibung eines Medikaments prognostiziert. Von isolierten Ärzten wurde angenommen, daß sie sich z.B. aus einer Fachzeitschrift über neue Medikamente informieren, d.h. von einer exogenen Quelle angeregt werden. Ist $dP(t)/dt$ der Zuwachs des Anteils von neu verschreibenden Ärzten, $P(t)$ der Anteil neu verschreibender Ärzte in der Population und c eine Konstante, so gilt für die isolierten Ärzte:

$$\frac{dP(t)}{dt} = c\ (1-P(t)).$$

Für die in das Kollegen-Netzwerk integrierten Ärzte wird hingegen der auf gegenseitigen Kontakten beruhende "logistische" Ausbreitungsprozeß vermutet:

$$\frac{dP(t)}{dt} = k\ P(t)\ (1-P(t)),$$

wobei k wiederum eine Konstante bezeichnet.

Der Zusammenhang mit der Ereignisdatenanalyse ist offensichtlich. Die Zeit bis zur Übernahme einer Neuerung ist die Ankunftszeit, $P(t)$ entspricht der kumulierten Wahrscheinlichkeitsverteilung $F(t)$, $1-P(t)$ der Überlebensfunktion $G(t)$und $dP(t)/dt$ der Dichteverteilung $f(t)$. Dann folgt aus den beiden Diffusionsmodellen für die Übergangsrate $r(t) = f(t)/(1-F(t))$ bei den isolierten Personen

$$r(t) = c$$

und für die integrierten Personen die monoton steigende Übergangsrate

$$r(t) = k\ F(t).$$

Das gewählte Diffusionsmodell impliziert somit eindeutig einen spezifischen Verlauf der Übergangsrate. Dies ist eine Konsequenz, die anhand von Daten über die Ausbreitung von Neuerungen prüfbar ist. Die Ereignisdatenanalyse erlaubt darüber hinaus die Schätzung der Parameter von Diffusionsmodellen unter Berücksichtigung von Kovariaten. Für isolierte "Neuerer" wäre somit das Exponentialmodell zu spezifizieren, wobei $r(t)=c$ von verschiedenen Merkmalen der Personen und ihres Kontextes abhängen könnte. Anders als bei herkömmlichen Verfahren können zensierte Zeiten wiederum angemessen berücksichtigt werden.

Eine Vielzahl von Diffusionsmodellen ergeben sich als Spezialfälle aus einer allgemeineren Diffusionsgleichung

$$\frac{dP(t)}{dt} = g(t)\ F(t)^m(1-F(t))^n$$

mit der Übergangsrate

$$r(t) = g(t)\ F(t)^m(1-F(t))^{n-1}$$

So ergibt sich z.B. für m=1, n=1 und g(t)=k das obige logistische Modell; für g(t)=a exp(-bt) das Hernes-Modell, für g(t)=a/t das in der Ereignisdatenanalyse wohlbekannte log-logistische Modell (Diekmann 1988), für n=2 und g(t)=c das Floyd-Modell (Mahajan/Peterson 1985) und für g(t)=c, m=0 und n=1 das Exponentialmodell.

Diffusionsmodelle sind somit spezielle parametrische Modelle der Ereignisdatenanalyse. Daraus ergeben sich eine Reihe von Konsequenzen für die empirische Analyse von Prozessen sozialer Diffusion. Die Ereignisdatenanalyse ermöglicht genauere Tests der Modelle sowie verbesserte Parameterschätzungen unter Berücksichtigung zensierter Daten. Durch die Einbeziehung von Kovariaten bei der Untersuchung von Diffusionsprozessen (Krempel 1987, Marsden/ Podolny 1988). kann der beobachteten Heterogenität realistischerweise Rechnung getragen werden.

Darüber hinaus kann das Spektrum der in der Ereignisdatenanalyse verwendeten Modelle durch die Einbeziehung von Diffusionsmodellen erweitert werden. Von Interesse ist hierbei auch, daß verschiedene Diffusionsmodelle nicht-monotone parametrische Verläufe der Übergangsrate implizieren (z.B. das Hernes-Modell und das Floyd-Modell). Als Alternative zu den "klassischen" Modellen eröffnen Diffusionsmodelle für die Ereignisdatenanalyse neue Möglichkeiten. Dies gilt um so mehr, da heutzutage mit dem Programm GAUSS die Likelihoodfunktionen neuer Modelle mit geringerem Aufwand als früher programmierbar sind. Der programmtechnische Aufwand für die Parameterschätzung wird dadurch wesentlich vermindert.

Es ist zweifellos eine interessante Forschungsperspektive, die mathematischen Eigenschaften von Diffusionsmodellen und ihrer jeweiligen Übergangsratenfunktionen genauer zu untersuchen, die Parameterschätzung softwaremäßig zu realisieren und praktische Anwendungen, z.B. bei der Ausbreitung neuer Techniken, in der Arbeitsmarktforschung und bei demographischen Prozessen zu erproben.

5. Probleme und Perspektiven

Wenn auch die Ereignisdatenanalyse zunehmend in das Standard-Repertoire der empirischen Sozialforschung eingeht, so besteht doch noch ein erheblicher Forschungsbedarf zur Verbesserung und Weiterentwicklung der Methoden und zur Klärung zahlreicher Anwendungsprobleme. Der Forschungsbedarf bezieht sich dabei auf vier Ebenen: die Modellkonstruktion und Weiterentwicklung von Modellen, Probleme der statistischen Schätzung, Probleme bei der Erhebung zeitbezogener Daten und die Bereitstellung geeigneter Software zur Datenanalyse. Eine Liste der Weiterentwicklungen und Forschungsprobleme wird u.a. die folgenden Aspekte umfassen:

Modelle mit kontinuierlicher Zustandsraumvariable (Tuma/Hannan 1984: Kap. 15), Modelle mit diskreter Zeit (Hamerle 1985), Modelle mit nicht-monotonen Ratenfunktionen (Diekmann/Mitter 1983, 1984b; Diekmann 1988), Mischverteilungsmodelle zur Berücksichtigung unbeobachteter Heterogenität (Heckmann/ Singer 1982; Trussel/Richards 1985; Vaupel/Yashin 1985) sowie Nicht-Markov-Modelle, die die Abhängigkeit der Übergangsrate von vorhergehenden Ereignissen und Verweildauern explizit berücksichtigen (Heckmann/Borjas 1980). Wie im vorhergehenden Abschnitt erläutert, ist es zudem von Interesse, das Potential von Diffusionsmodellen für die Ereignisdatenanalyse genauer zu erforschen.

Bei der statistischen Schätzung können u.a. die folgenden Probleme auftreten: Links-zensierte Daten und Sample-Selektions-Probleme in Abhängigkeit von gewählten Erhebungsdesigns (Hamerle 1986), Verzerrungen durch Spezifikationsfehler, Meßfehler bei Ankunftszeiten und Kovariaten (Arminger 1984; Galler 1985; Tuma/Hannan 1984:145ff.), interne zeitabhängige Kovariate (Kalbfleisch/Prentice 1980) sowie korrelierte Risiken und Zensierungsmechanismen. Von Interesse ist hier auch die Erprobung von Spezifikationstests (Arminger 1987; Hausman 1978; White 1981) und die Durchführung von Simulationsstudien zur Untersuchung der Robustheit und small-sample-Eigenschaften der Schätzer (Tuma/Hannan 1984: Kap. 5). Bootstrap-Techniken (Rothe 1986, 1988), die in der Ereignisdatenanalyse bislang wenig benutzt wurden, könnten dabei wertvolle Aufschlüsse geben.

Auch die ausgeklügelsten statistischen Analyseverfahren führen zu verzerrten Ergebnissen, wenn die erhobenen Daten wenig verläßlich sind. In zahlreichen Untersuchungen werden Ereignisdaten retrospektiv erfragt, neuerdings auch mit telefonischen Interviewtechniken. Andere Untersuchungen greifen dagegen auf prozeß-produzierte Ereignisdaten zurück (z.B. Daten aus der Buchhaltung einer Firma, Sozialversicherungsdaten etc.). Es fragt sich aber, wie valide die Daten bei welchen erhobenen Merkmalen und Erhebungsverfahren sind und wie sich diese gegebenenfalls verbessern lassen (Tölke 1980; Preisendörfer 1987). Vielfach werden auch nicht die exakten Ankunftszeiten vorliegen, sondern nur die Häufigkeit von Ereignissen ("event counts") oder aber Ereignisabfolgen ("event sequence data") bekannt sein. Von Bedeutung sind daher Untersuchungstechniken, die die Schätzung von Übergangsratenmodellen auch dann erlauben, wenn weniger informationshaltige Daten als vollständige "event-histories" vorliegen (Coleman 1981).

Weniger informationshaltig als vollständige Ereignisdaten sind auch sogenannte Quantal-Response-Daten ("inspection data", Intervalldaten, vgl. Nelson 1982:405-432). Bei diesen Daten ist nur bekannt, ob ein bestimmtes Ereignis in einem Zeitabschnitt der Länge t aufgetreten ist (d=1, andernfalls d=0). Unbekannt hingegen bleibt die exakte Länge der Ankunftszeit. Obgleich eine derartige Datenstruktur häufig gegeben ist, wurden Quantal-Response-Daten in den Sozialwissenschaften nur selten zur Schätzung von Übergangsratenmodellen herangezogen. Wird ein parametrisches Modell spezifiziert, so kann für den Likelihoodbeitrag der Person i der folgende Ausdruck geschrieben werden:

$$L_i = F(t_i)^{d_i} \; G(t_i)^{(1-d_i)}.$$

Anstelle der Dichtefunktion wird für die nicht-zensierten Zeiten die kumulierte Verteilungsfunktion in die Likelihood-Funktion eingesetzt. Ein interessanter Anwendungsfall sind die Angaben zur Ehescheidung im amerikanischen "General Social Survey". Gefragt wird im GSS nach dem Heiratsalter einer Person sowie danach, ob die Person jemals geschieden wurde. Nicht erfragt wurde jedoch das Scheidungsdatum, so daß keine vollständigen Ereignisdaten vorliegen. Dennoch dürfte es in der beschriebenen Weise möglich sein, Übergangsratenmodelle anhand der GSS-Scheidungsdaten unter Berücksichtigung von Kovariaten zu schätzen.

Verschiedene Neuentwicklungen sind ferner im Bereich der Software zur Analyse von Ereignisdaten zu verzeichnen. Eine Vielzahl parametrischer Modelle kann heute u.a. mit den Programmen RATE, GLIM (Arminger 1986), BMDP.P3RFUN (Petersen 1986; Brüderl/Wallascheck 1988), LIMDEP, SAS und GAUSS-PARAT (Schneider 1988) geschätzt werden. Ein geplanter Workshop im kommenden Frühjahr bei ZUMA soll speziell der Softwareentwicklung zur Analyse von Ereignisdaten gewidmet sein.

Dieser Beitrag wurde von Andreas Diekmann verfaßt.

Literatur

Allison, P.D., 1984: Event History Analysis. Beverly Hills/London/New Dehli: Sage.
Andress, H.J., 1985: Multivariate Analyse von Verlaufsdaten. Mannheim: ZUMA.
Arminger, G., 1984: Mathematische und methodische Probleme bei der Analyse von Paneldaten mit qualitativen Variablen. Vierteljahreshefte des deutschen Instituts für Wirtschaftsforschung, S.470-480.
Arminger, G., 1986: Analysis of event histories with generalized linear models. S.245-283 in: A. Diekmann/P. Mitter, 1986.
Arminger, G., 1987: Testing against misspecification in parametric rate models. S.679-699 in: K.U. Mayer/N.B. Tuma.
Blossfeld, H.P./Hamerle, A./Mayer, K.U., 1986: Ereignisanalyse. Statistische Theorie und Anwendungen in den Wirtschafts- und Sozialwissenschaften. Frankfurt: Campus.
Brüderl, J./Wallascheck, 1988: Beschreibung eines BMDP-Programms zur Schätzung parametrischer Survival-Modelle. Unveröffentlichtes Manuskript: Sonderforschungsbereich 333 der Universität München.
Coleman, J.S., 1981: Longitudinal Data Analysis. New York: Basic Books.

Coleman, J.S./Katz, E./Menzel, H., 1957: The diffusion of an innovation among physicians. Sociometry 20:253-270.

Diekmann, A., 1987: Determinanten des Heiratsalters und Scheidungsrisikos. Unveröffentlichte Habilitationsschrift, Fakultät für Sozialwissenschaften der Universität München.

Diekmann, A., 1988: Diffusion and survival models for the entry into marriage. Journal of Mathematical Sociology 14 (im Druck).

Diekmann, A./Mitter, P., 1983: The 'sickle hypothesis'. A time dependent Poisson model with applications to deviant behavior and occupational mobility. Journal of Mathematical Sociology 9:85-101.

Diekmann, A./Mitter, P., 1984a: Methoden zur Analyse von Zeitverläufen. Stuttgart: Teubner.

Diekmann, A./Mitter, P. (Hrsg.), 1984b: Stochastic Modelling of Social Processes. Orlando: Academic Press.

Diekmann, A./Mitter, P. 1984c: A comparison of the 'sickle function' with alternative stochastic models of divorce rates. S. 123-153 in: Diekmann/Mitter 1984b.

Diekmann, A./Preisendörfer, P., 1988: Turnover and employment stability in a large West-German company. European Sociological Review 4.

Elandt-Johnson, R.C./Johnson, N.L., 1980: Survival Models and Data Analysis. New York: Wiley.

Galler, H.P., 1985: Übergangsratenmodelle bei intervalldatierten Ereignissen. Arbeitspapier Nr. 164 des SFB 3 der Universitäten Frankfurt und Mannheim.

Hamblin, R.L./Jacobsen, R.B./Miller, J.L.L., 1973: A Mathematical Theory of Social Change. New York: Wiley.

Hamerle, A., 1985: Diskrete Modelle zur statistischen Analyse von Verweildauern und Lebenszeiten. Diskussionsbeiträge der Fakultät für Wirtschaftswissenschaften und Statistik Nr. 79. Universität Konstanz.

Hamerle, A., 1986: Statistische Analyse links-zensierter Daten und spezieller Erhebungsdesigns in Duration-Modellen. Diskussionsbeiträge der Fakultät für Wirtschaftswissenschaften und Statistik Nr.93. Universität Konstanz.

Hausman, J.A., 1978: Specification Tests in Econometrics. Econometrica 46:1251-1272.

Heckman, J.J./Borjas, G.J., 1980: Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence. Economica 47:247-283.

Heckman, J.J./Singer, B., 1982: Population heterogeneity in demographic models. S.567-599 in: K.C. Land/A. Rogers (Hrsg.), Multidimensional Mathematical Demography. New York: Academic Press.

Heckman, J.J./Singer, B., 1984a: Econometric duration analysis. Journal of Econometrics 24:63-132.

Heckman, J.J./Singer, B., 1984b: A method for minimizing the impact of distributional assumptions in econometric models for duration data. Econometrica 52:271-320.

Heinritz, S./Jungbauer-Gans, M./Krassnig, H., 1986: Verschiedene Methoden zur Analyse von Zeitverläufen dargelegt am Beispiel von Aggression im Straßenverkehr. Seminararbeit am Institut für Soziologie der Universität München.

Kalbfleisch, J.D./Prentice, R.L., 1980: The Statistical Analysis of Failure Time Data. New York: Wiley.

Krempel, L., 1987: Soziale Interaktionen: Einstellungen, Biographien, Situationen und Beziehungsnetzwerke. Dynamische Ereignisanalysen. Bochum: Schallwig Verlag.

Lawless, J.F. 1982: Statistical Models and Methods for Life-Time Data. New York.

Mahajan, V./Peterson, R.A., 1985: Models for Innovation Diffusion. Beverly Hills/London/New Dehli: Sage.

Marsden, P.V./Podolny, J., 1988: Dynamic Analysis of Network Diffusion Processes. MASO-Conference on Network Analysis, June 8-11, Utrecht.

Mayer, K.U./Tuma, N.B. (Hrsg.), 1987: Applications of Event History Analysis in Life Course Research, Materialien aus der Bildungsforschung Nr. 30. Berlin: Max-Planck-Institut für Bildungsforschung.

Nelson, W., 1982: Applied Life Data Analysis. New York: Wiley.

Peterson, T., 1986: Estimating fully parametric hazard rate models with time-dependent covariates. Use of maximum likelihood. Sociological Methods and Research 14:219-246.

Preisendörfer, P., 1987: "Life Histories": Neuere Verfahren zur Sammlung retrospektiver Daten, insbesondere Berufsverlaufsdaten. Arbeitspapier des Sonderforschungsbereichs 333 der Universität München.

Rothe, G., 1986: Some remarks on bootstrap techniques for constructing confidence intervals. Statistical Papers 27:165-172.

Rothe, G., 1988: Jackknife und Bootstrap: Ein Überblick. ZUMA-Arbeitsbericht (in Vorbereitung).

Schneider, H., 1988: PARAT. Benutzer-Handbuch - Version 1.01. Arbeitspapier des Sonderforschungsbereichs 3 der Universitäten Frankfurt und Mannheim.

Tölke, A., 1980: Zuverlässigkeit retrospektiver Verlaufsdaten. Qualitative Ergebnisse einer Nachbefragung. Arbeitspapier Nr. 30 des Sonderforschungsbereichs 3 der Universitäten Frankfurt und Mannheim.

Trussel, J./Richards, T., 1985: Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer-procedure. S.242-276 in: Tuma, N.B. (Hrsg.), Sociological Methodology. San Francisco: Jossey Bass.
Tuma, N.B./Hannan, M., 1984: Social Dynamics. Models and Methods. Orlando: Academic Press.
Vaupel, J.W./Yashin, A.I., 1985: The deviant dynamics of death in heterogeneous populations. S.179-211 in: N.B. Tuma (Hrsg.), Sociological Methodology 1985. San Francisco: Jossey Bass.
White, H., 1981: Consequences and detection of misspecified nonlinear regression models. Journal of the American Statistical Association 75:419-433.